A comprehensive framework for Controlled Query Evaluation, Consistent Query Answering and KB Updates in Description Logics

Domenico Lembo, Riccardo Rosati, Domenico Fabio Savo
Sapienza University of Rome
lastname@diag.uniroma1.it

In this extended abstract we discuss the relationship between confidentiality-preserving frameworks and inconsistency-tolerant repair and update semantics in Description Logics (DL). In particular, we consider the well-known problems of Consistent Query Answering, Controlled Query Evaluation, and Knowledge Base Update in DL and introduce a unifying framework that can be naturally instantiated to capture significant settings for the above problems, previously investigated in the literature.

Consistent Query Answering (CQA). It is a declarative approach to manage inconsistency that has been extensively studied in both databases and knowledge bases (KBs) (Bertossi 2011; Bienvenu and Bourgaux 2016). The crucial notions for CQA are those of *repair* and *consistent answers*. For a DL KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, where $\mathcal{T}$ is a consistent TBox and $\mathcal{A}$ is an ABox, a repair is an ABox $\mathcal{A}'$ that "minimally" differs from $\mathcal{A}$ and that is consistent with $\mathcal{T}$. Several notions of minimality can be adopted, which in general give rise to the existence of several repairs. For example, in the *AR*-semantics, a repair is a maximal subset of $\mathcal{A}$ that is consistent with $\mathcal{T}$ (Lembo et al. 2010). In the *CAR*-semantics, a repair is a maximal subset of the consistent closure of $\mathcal{K}$ that is consistent with $\mathcal{T}$, where the consistent closure of $\mathcal{K}$ is the set of facts implied by $\mathcal{T}$ and by any subset of $\mathcal{A}$ consistent with $\mathcal{T}$ (Lembo et al. 2010). The consistent answers to a query q specified over $\mathcal{K}$ are the answers to q that are true in every $\mathcal{K}' = \langle \mathcal{T}, \mathcal{A}' \rangle$, where $\mathcal{A}'$ is a repair of $\mathcal{A}$.

Controlled Query Evaluation (CQE). It is a confidentiality-preserving framework, in which a certain *policy* is declaratively specified on the schema level of a database or knowledge base to regulate the access to the underlying data. Users ask queries on the schema, but when these are answered no secrets have to be disclosed. To this aim, a so-called *censor* can alter the returned answers according to the specified policy. The main reasoning task in CQE thus amounts to computing an *optimal* censor, given a KB or a database, a policy, and a query language. Intuitively, to be optimal a censor has to guarantee the policy by "minimally" modifying the answers to queries, so that it hides only the portion of data that is necessary to preserve secrets, but at the same time it returns as many answers as possible among those that the system is allowed to return. Various ways to obtain a censor have been investigated, which at an abstract level correspond to various criteria of minimal modification of the query answers. The notion of censor has been first studied in the context of databases (e.g., (Biskup and Bonatti 2004)), but recently has been investigated in the setting of DL ontologies. In this paper we mainly refer to the framework studied in (Cuenca Grau et al. 2015), where the policy is defined as a conjunctive query over a DL KB, with the intended meaning that the answers to this query should not be returned to the users.

Knowledge Base Update (KBU). It is the problem of modifying a KB in order to adhere to a change in the domain of interest. Typically such a change is represented by means of a set of formulas that have to be entailed by the KB resulting from the update. If the new knowledge contradicts the old one, a portion of the latter has to be discarded for restoring the consistency. Similarly to CQA, the idea is to "minimize the distance" between the original KB and the new one resulting from the application of an evolution operator implementing the update (see (Flouris et al. 2008) for a survey). In our investigation, we focus on the so-called formula-based approaches, which define the result of the update in terms of a formula that differs minimally from the formula expressing the original KB. Basically, solutions for this case can be divided into two categories: foundational approaches and coherence approaches (Gärdenfors 1990). The two differ since in the foundational approaches the axioms asserted in the KB serve as a justification for all the knowledge of the KB, while in the coherence approaches axioms are seen only as a means of expressing the knowledge. Intuitively, given two KBs $\mathcal{K}$ and $\mathcal{K}_{\mathcal{N}}$, for updating $\mathcal{K}$ with $\mathcal{K}_{\mathcal{N}}$ in the foundational approaches the maximal subsets of $\mathcal{K}$ that are consistent with $\mathcal{K}_{\mathcal{N}}$ need to be computed. Instead, in the coherence approaches, one intends to preserve as many assertions as possible of those entailed by $\mathcal{K}$, and thus aims to compute the maximal subsets of the logical consequences of $\mathcal{K}$ that are consistent with $\mathcal{K}_{\mathcal{N}}$.

Motivations. The three problems we are considering can be actually seen as *belief revision* problems. Belief revision studies the issue of integrating new information with previous knowledge. Whereas the new knowledge has to be preserved in the resulting theory, the old one should be revised to guarantee consistency. In CQA in DL the TBox can be seen as the new information reflecting a modification about the knowledge of the domain, whereas the previous knowledge is the ABox. In KBU the new knowledge is explicit in the framework and reflects a change in the world. In CQE in DL the policy together with the TBox of the KB can be considered as the new knowledge, whereas the underlying ABox plays the role of the previous knowledge. Under this perspective, for instance, the policy considered in (Cuenca Grau et al. 2015) is seen as a constraint over the KB obtained by negating the policy query, and confidential data are those facts in the ABox that violate such constraint. As belief revision problems, CQA, CQE and KBU share the objective of modifying knowledge to ensure consistency by applying minimal changes to the original theory. In CQE, for instance, this means to minimally alter query answers to enforce the policy.

Framework. Based on the above common aspects shared by CQA, CQE and KBU, we introduce below the new notion of *consistent projections of a KB*. Intuitively, the consistent projection of a KB $\mathcal{K}$ is a KB that maximally preserves the knowledge that $\mathcal{K}$ infers in a certain projection language $\mathcal{Q}$ and that is consistent with some new knowledge, which we denote with $\mathcal{K}_c$. In the following we indicate with $Mod(\mathcal{K})$ the set of models of $\mathcal{K}$, and with $\mathcal{K} \models \phi$ the fact that $\mathcal{K}$ implies a first-order sentence ϕ. We also recall that $\mathcal{K}$ is consistent if $Mod(\mathcal{K}) \neq \emptyset$, inconsistent otherwise.

We give now the notion of *consistent $\mathcal{Q}$-entailment set of* $\mathcal{K}$, denoted by $CES(\mathcal{K}, \mathcal{Q})$, which is the set of sentences

$$\{\phi \mid \phi \in \mathcal{Q} \text{ and there exists } \mathcal{K}' \subseteq \mathcal{K} \text{ such that } Mod(\mathcal{K}') \neq \emptyset \text{ and } \mathcal{K}' \models \phi\}.$$

If $\mathcal{K}$ is a consistent KB, $CES(\mathcal{K}, \mathcal{Q})$ contains all the sentences in the language $\mathcal{Q}$ that are implied by $\mathcal{K}$. If $\mathcal{K}$ is inconsistent, $CES(\mathcal{K}, \mathcal{Q})$ resorts to consistent subsets of $\mathcal{K}$, and contains all the sentences in $\mathcal{Q}$ that are implied by at least one of these subsets.

We are now ready to formally define the notion of consistent projection of a KB.

Definition 1 Given a KB $\mathcal{K}$, a consistent KB $\mathcal{K}_c$, and a language $\mathcal{Q}$ of first-order sentences, a *$\mathcal{K}_c$-consistent $\mathcal{Q}$-projection of* $\mathcal{K}$ is a set of sentences Φ such that:

1. $\Phi \subseteq CES(\mathcal{K}, \mathcal{Q})$
2. $Mod(\mathcal{K}_c \cup \Phi) \neq \emptyset$
3. there exists no Φ' such that $\Phi \subset \Phi' \subseteq CES(\mathcal{K}, \mathcal{Q})$ and $Mod(\mathcal{K}_c \cup \Phi') \neq \emptyset$.

The set of the $\mathcal{K}_c$-consistent $\mathcal{Q}$-projections of $\mathcal{K}$ is denoted by $CP_{\mathcal{Q}}(\mathcal{K}, \mathcal{K}_c)$.

Therefore, a $\mathcal{K}_c$-consistent $\mathcal{Q}$-projection of $\mathcal{K}$ is a maximal subset of the consistent $\mathcal{Q}$-entailment set of $\mathcal{K}$ that is consistent with $\mathcal{K}_c$. Intuitively, with the above notion we aim to capture all the logical consequences of all the consistent portions of $\mathcal{K}$ that are expressible in $\mathcal{Q}$ (i.e., the $\mathcal{Q}$-projections of $\mathcal{K}$) and that are consistent with $\mathcal{K}_c$.

The above framework can be instantiated to express relevant CQA, CQE, and KBU settings studied in the literature. For example, it can be shown that the set of repairs under the *AR*-semantics for CQA over a DL KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ corresponds to the set $CP_{\mathcal{Q}}(\mathcal{K}, \mathcal{K}_c)$, when $\mathcal{K}_c = \mathcal{T}$ and $\mathcal{Q} = \mathcal{A}$, i.e., if the projection language includes only the facts of the ABox. By suitably setting $\mathcal{K}_c$ and $\mathcal{Q}$ we can also show that the notion of consistent projection captures the one of repair under the *CAR*-semantics studied in CQA for DL, the one of optimal censor proposed in (Cuenca Grau et al. 2015), and general notions of update in the foundational and coherence formula-based approaches and their instance-level versions for DL studied in (De Giacomo et al. 2016).

In our framework we study a form of entailment that captures and generalizes skeptical entailment in CQA, CQE and KBU under the correspondences mentioned above. A formal definition is given below.

Definition 2 Given a KB $\mathcal{K}$, a consistent KB $\mathcal{K}_c$, a KB $\mathcal{K}_n \subseteq \mathcal{K}_c$, a projection language $\mathcal{Q}$ and a first-order sentence ϕ, *Consistent Projection Entailment* is the problem of deciding whether $\mathcal{K}_n \cup \Phi \models \phi$ for every $\Phi \in CP_{\mathcal{Q}}(\mathcal{K}, \mathcal{K}_c)$.

In the above definition, $\mathcal{K}_n$ is the portion of $\mathcal{K}_c$ that is relevant for reasoning. To have an intuition on the role of $\mathcal{K}_n$, consider for example CQE under the perspective of our framework, where consistent projections are computed with respect to the DL TBox and (the negation of) the policy, whereas query answering is on the TBox only (and the underlying data).

We have some preliminary complexity results for this general form of entailment that straightforwardly apply to CQA, CQE and KBU, and we believe that our framework paves the way for a unified approach to such problems.

References

Bertossi, L. E. 2011. *Database Repairing and Consistent Query Answering*. Synthesis Lectures on Data Management. Morgan & Claypool Publishers.

Bienvenu, M., and Bourgaux, C. 2016. Inconsistency-tolerant querying of description logic knowledge bases. In *RW 2012 Tutorial Lectures*, 156–202.

Biskup, J., and Bonatti, P. A. 2004. Controlled query evaluation for enforcing confidentiality in complete information systems. *Int. J. Inf. Sec.* 3(1):14–27.

Cuenca Grau, B.; Kharlamov, E.; Kostylev, E. V.; and Zheleznyakov, D. 2015. Controlled query evaluation for datalog and OWL 2 profile ontologies. In *Proc. of IJCAI 2015*, 2883–2889.

De Giacomo, G.; Oriol, X.; Rosati, R.; and Savo, D. F. 2016. Updating DL-Lite ontologies through first-order queries. In *Proc. of ISWC 2016*, 167–183.

Flouris, G.; Manakanatas, D.; Kondylakis, H.; Plexousakis, D.; and Antoniou, G. 2008. Ontology change: Classification and survey. *Knowledge Engineering Review* 23(2):117–152.

Gärdenfors, P. 1990. The dynamics of belief systems: Foundations vs. coherence theories. *Revue internationale de philosophie* 24–46.

Lembo, D.; Lenzerini, M.; Rosati, R.; Ruzzi, M.; and Savo, D. F. 2010. Inconsistency-tolerant semantics for description logics. In *Proc. of RR 2010*, 103–117.